JOURNAL OF APPROXIMATION THEORY 18, 264–277 (1976)

An Overdetermined Linear System

MICHAEL L. LEVITAN AND ROGER Y. S. LYNN

Department of Mathematics, Villanova University, Villanova, Pennsylvania 19085

Communicated by E. W. Cheney

Received April 4, 1975

1. Introduction

Consider a linear system of algebraic equations $Ax = b$, where $A = (a_{ij})$ is an $(n+1) \times n$ real matrix, $b = (b_i)$ an $(n+1)$-component real vector and $x = (x_i)$ an n-component real vector. Clearly, the system is overdetermined if b is not in the range of A and it is this problem to which we seek a "best" approximate solution. The technique employed is to impose an abstract norm on R^{n+1} and then find $x \in R^n$ such that the error vector $\eta(x) = b - Ax$ is minimized with respect to this norm. We refer to this question of finding such a solution with respect to a given abstract norm as "Problem (P)." When we use the standard l^p-norm, we let $\xi_p = (\xi_i{}^p)$ denote an l^p-solution to $Ax = b$, i.e.,

$$\| \eta(\xi_p)\|_p = \min_{x \in R^n} \| \eta(x)\|_p .$$

(The least-squares solution and the Tchebychev solution correspond to the values $p = 2$ and $p = \infty$, respectively.)

A set of vectors in R^m is said to satisfy the Haar condition if every set of m of them is linearly independent. Let $\mathscr{H}$ denote the set of all $(n+1) \times n$ matrices whose rows satisfy the Haar condition. In this paper, we concern ourselves with the problem of finding the l^p-solution to $Ax = b$ when $A \in \mathscr{H}$. In Section 2, we give an explicit formula for the l^p-solution and also discuss how it may be expressed as a convex combination of the solutions to the $n \times n$ subsystems. In Section 3, we allow b to be a random vector and demonstrate how one may favor a particular norm when observing a minimum variance criterion. Two examples of such stochastic overdetermined systems are given. Finally, some preliminary results pertaining to the matrix A are presented in the Appendix.

2. Explicit Forms of the l^p-Solution

Let A be an $(n+1) \times n$ matrix of rank n and A^{T} the transpose of A. Then $Ax = b$ has the unique l^2-solution

$$\xi_2 = (A^{\mathrm{T}}A)^{-1} A^{\mathrm{T}}b.$$

Moreover, the l^2-error vector is $s = \eta(\xi_2) = (I - A(A^{\mathrm{T}}A)^{-1} A^{\mathrm{T}})\, b$ (see [1]).

The dual norm $\|\cdot\|^{\sim}$ of a norm $\|\cdot\|$ on R^{n+1} is defined by

$$\| u \|^{\sim} = \max_{\|v\|=1} (u, v),$$

where

$$(u, v) = \sum_{i=1}^{n+1} u_i v_i$$

is the standard Euclidean inner product on R^{n+1}. Furthermore, $\tilde{u}$ is called a dual vector to $u \in R^{n+1}$ if $(\tilde{u}, u) = \| u \|^{\sim}$ and $\| \tilde{u} \| = 1$. From [4] we have the following

THEOREM (Sreedharan). *Let A be an $(n+1) \times n$ matrix of rank n and $s = \eta(\xi_2) = b - A\xi_2$ the l^2-error vector. If $s = 0$, then ξ_2 is a solution of Problem (P). If $s \neq 0$, then*

$$Ax = b - ((b, s)/\| s \|^{\sim})\, \tilde{s} \tag{2.1}$$

has a solution, and any solution of (2.1) *is a solution of Problem (P).*

From this point we shall only consider the case where R^{n+1} is equipped with the l^p-norm. Thus, the l^q-norm is the dual norm if $(1/p) + (1/q) = 1$. For $1 < p \leqslant \infty$, the dual vector of a nonzero vector $s = (s_i)$ is $\tilde{s}_p = (\tilde{s}_i{}^p)$, where

$$\tilde{s}_i{}^p = (| s_i |/\| s \|_q)^{q-1} \operatorname{sgn} s_i, \qquad i = 1, 2, \ldots, n+1. \tag{2.2}$$

For $p = 1$, we define $\rho = \{r_1, r_2, \ldots, r_t\} = \{r\colon 1 \leqslant r \leqslant n+1$ and $| s_r | = \max_i | s_i | = \| s \|_\infty\}$ and $\tilde{s}_1(r) = (\tilde{s}_i{}^1(r))$, where

$$\begin{aligned} \tilde{s}_i{}^1(r) &= \operatorname{sgn} s_r \qquad \text{if} \quad i = r, \\ &= 0 \qquad\qquad \text{if} \quad i \neq r. \end{aligned}$$

Then any convex combination of $\tilde{s}_1(r)$ with $r \in \rho$ is a dual vector of s for $p = 1$.

Clearly

$$\lim_{p\to 1^+} \tilde{s}_p = (1/t) \sum_{j=1}^{t} \tilde{s}_1(r_j) \tag{2.3}$$

is also a dual vector of s for $p = 1$.

With the given l^p-norm and $s \neq 0$, we may now write (2.1) as $Ax = b^{(p)}$ where

$$b^{(p)} = (b_i^{(p)}) = b - ((b, s)/\| s \|_q) \tilde{s}_p \tag{2.4}$$

and $\tilde{s}_p$ is given by (2.2) for $1 < p \leqslant \infty$, while any convex combination of $\tilde{s}_1(r)$ with $r \in \rho$ can define $\tilde{s}_1$. It is shown in [1, 2] that $Ax = b^{(p)}$ has a unique solution for each p, $1 \leqslant p \leqslant \infty$. Furthermore, each is a respective l^p-solution to $Ax = b$. The l^p-solution to $Ax = b$ is unique when $1 < p < \infty$ since the l^p-norm is strictly convex. From [2] we have that the l^1-solution is unique if and only if ρ is a singleton set ($t = 1$), and the l^∞-soution is unique if and only if the rows of A satisfy the Haar condition. If $Ax = b^{(p)}$ has a unique solution, then it can be found by solving $A^T Ax = A^T b^{(p)}$. Hence the l^p-solution can be obtained from

$$\xi_p = (A^T A)^{-1} A^T b^{(p)}. \tag{2.5}$$

Let A^k be the $n \times n$ matrix obtained from A by deleting the kth row, $D_k \equiv \det(A^k)$, the determinant of A^k,

$$\Delta_q \equiv \sum_{k=1}^{n+1} | D_k |^q$$

and

$$\sigma \equiv \sum_{k=1}^{n+1} (-1)^{k-1} b_k D_k .$$

Now $A \in \mathscr{H}$ implies that there exists a nonsingular $n \times n$ matrix P with $| \det(P)| = 1$ such that $AP = G$ where

$$G = \begin{pmatrix} g_1 & 0 & \cdots & 0 \\ 0 & g_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & g_n \\ h_1 & h_2 & \cdots & h_n \end{pmatrix}$$

with $g_i \neq 0$ and $h_i \neq 0$, $i = 1, 2, \ldots, n$. Let G^k be the matrix G with the kth row deleted. Then for $k = 1, 2, \ldots, n + 1$

$$A^k P = G^k$$

and

$$\hat{D}_k \equiv \det(G^k) = D_k \det(P).$$

Consider the $(n+1) \times (n+1)$ matrix formed by extending A in the following way:

$$(\alpha_{ij}) \equiv (A \mid \delta),$$

where $\delta \equiv ((-1)^{n+1-i} D_i)$ is an $(n+1)$-component column vector. Define $m_{ij} = (\text{cofactor of } \alpha_{ji})/\Delta_2$ for $i = 1,\ldots, n, j = 1,\ldots, n+1$, or more explicitly,

$$m_{ij} = \frac{(-1)^{i+j}}{\Delta_2} \begin{vmatrix} a_{11} & \cdots & a_{1,i-1} & a_{1,i+1} & \cdots & a_{1n} & (-1)^n D_1 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_{j-1,1} & \cdots & a_{j-1,i-1} & a_{j-1,i+1} & \cdots & a_{j-1,n} & (-1)^{n-j+2} D_{j-1} \\ a_{j+1,1} & \cdots & a_{j+1,i-1} & a_{j+1,i+1} & \cdots & a_{j+1,n} & (-1)^{n-j} D_{j+1} \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ a_{n+1,1} & \cdots & a_{n+1,i-1} & a_{n+1,i+1} & \cdots & a_{n+1,n} & (-1)^0 D_{n+1} \end{vmatrix}. \tag{2.6}$$

THEOREM 2.1. *Let $A \in \mathscr{H}$.*

(i) *If $\sigma = 0$, then $Ax = b$ is solvable and has the unique solution $x = Mb$ where $M = (m_{ij})$ is defined by* (2.6).

(ii) *If $\sigma \neq 0$, then an l^p-solution to $Ax = b$ is $\xi_p = B_p b$, where $B_p = (\beta_{ij}(p))$ is defined by*

$$\beta_{ij}(p) = m_{ij} + \frac{(-1)^j D_j}{\Delta_q} \sum_{k=1}^{n+1} (-1)^{k-1} \frac{m_{ik} \mid D_k \mid^q}{D_k}, \qquad 1 < p \leqslant \infty,$$

$$= m_{ij} + \frac{(-1)^j D_j}{\tau} \sum_{k\in\lambda} (-1)^{k-1} \frac{m_{ik}}{D_k}, \qquad p = 1,$$

$(1/p) + (1/q) = 1$, $\lambda = \{l_1, l_2,\ldots, l_\tau\} = \{l\colon 1 \leqslant l \leqslant n+1$ and $\mid D_l \mid = \max_i \mid D_i \mid\}$, and m_{ij} is given by (2.6). *The l^1-solution given here has the property that $\lim_{p\to1^+} \xi_p = \xi_1$, hence ξ_p is continuous with respect to p for $1 \leqslant p \leqslant \infty$. Furthermore, ξ_p is the unique l^p-solution for $1 < p \leqslant \infty$ while the l^1-solution is unique if and only if λ is a singleton set ($\tau = 1$).*

Proof. Since $s = (I - A(A^TA)^{-1} A^T)\, b$, Theorem A.1 (cf. Appendix) implies that

$$s_i = (\sigma/\Delta_2)(-1)^{i-1} D_i, \qquad i = 1, 2,\ldots, n+1. \tag{2.7}$$

(i) $A \in \mathscr{H}$ implies that rank $(A) = n$ and thus A is one-to-one. Now b is in the range of A if and only if $s = 0$ which, in turn, is equivalent to $\sigma = 0$.

Hence $Ax = b$ is uniquely solvable in this case and it has the solution $x = (A^{T}A)^{-1} A^{T}b = Mb$.

(ii) For $\sigma \neq 0$, we see from (2.7) that s_i is proportional to D_i, so ρ and λ represent the same set. Furthermore,

$$(b, s) = \sigma^2/\Delta_2 . \tag{2.8}$$

From (2.2), (2.3), (2.7), (2.8), and (2.4) we obtain

$$b_j^{(p)} = b_j + \frac{\sigma}{\Delta_q} \frac{(-1)^j \mid D_j \mid^q}{D_j}, \qquad 1 < p \leqslant \infty; \tag{2.9a}$$

$$\begin{aligned} b_j^{(1)} &= b_j + (\sigma(-1)^j/\tau D_j), && j \in \lambda, \\ &= b_j && j \notin \lambda. \end{aligned} \tag{2.9b}$$

Theorem A.2 (cf. Appendix) and (2.5) yield

$$\xi_i{}^p = \sum_{j=1}^{n+1} m_{ij} b_j^{(p)}, \qquad i = 1, 2,..., n; \qquad 1 \leqslant p \leqslant \infty;$$

hence

$$\xi_i{}^p = \sum_{j=1}^{n+1} \beta_{ij}(p)\, b_j , \qquad i = 1, 2,..., n, \qquad 1 \leqslant p \leqslant \infty,$$

via (2.9) and the definition of σ. We see that $b_j^{(p)}$, $\beta_{ij}(p)$, and thus ξ_p are continuous with respect to p for $1 \leqslant p \leqslant \infty$ when $\tilde{s}_1$ is defined by (2.3).

COROLLARY 2.1. *Let $A \in \mathscr{H}$, $\sigma \neq 0$, and $\mid D_i \mid = c$, $i = 1, 2,..., n + 1$, for some c. Then ξ_p is independent of p for $1 \leqslant p \leqslant \infty$.*

Proof. This follows immediately from the definition of $\beta_{ij}(p)$.

Let us now consider the $n \times n$ subsystem of equations

$$A^k Z^k = b^k \tag{2.10}$$

for $k = 1, 2,..., n + 1$, where b^k is the vector b with the kth element deleted. Then (2.10) has a unique solution Z^k since A^k is nonsingular ($D_k \neq 0$) by the Haar condition. We would like to establish a relation between the l^p-solution ξ_p and the Z^k's. For each $k = 1, 2,..., n + 1$, the $n \times n$ system

$$G^k W^k = b^k \tag{2.11}$$

also has a unique solution $W^k = (w_i{}^k)$ since $\mid \hat{D}_k \mid = \mid D_k \mid \neq 0$.

We note that

$$G^k = \begin{bmatrix} g_1 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & g_{k-1} & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & g_{k+1} & \cdots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & & 0 & 0 & 0 & \cdots & g_n \\ h_1 & \cdots & h_{k-1} & h_k & h_{k+1} & \cdots & h_n \end{bmatrix}.$$

Hence the first $n - 1$ equations of the system (2.11) yield

$$w_i^k = b_i/g_i \qquad \text{for} \quad i \neq k. \tag{2.12a}$$

By Cramer's rule and through expanding the determinant in the numerator about the kth column, we obtain

$$w_k^k = \frac{1}{\hat{D}_k} \sum_{j\neq k} (-1)^{k+j}\, b_j \left(-\frac{\hat{D}_j}{g_k}\right) = \frac{1}{D_k g_k} \sum_{j\neq k} (-1)^{k+j-1}\, b_j D_j\,. \tag{2.12b}$$

Lemma 2.1. (i) *If* $\sigma = 0$ *then* $(G^{\mathrm{T}}G)^{-1}\, G^{\mathrm{T}} b = W \equiv (b_i/g_i)$.

(ii) *If* $\sigma \neq 0$ *then*

$$\begin{aligned} (G^{\mathrm{T}}G)^{-1}\, G^{\mathrm{T}} b^{(p)} &= (1/\Delta_q) \sum_{k=1}^{n+1} |\, D_k\,|^q\, W^k, \qquad 1 < p \leqslant \infty, \\ &= (1/\tau) \sum_{k\in\lambda} W^k, \qquad\qquad p = 1, \end{aligned}$$

where $b^{(p)}$, λ *and* τ *were defined in Theorem* 2.1.

Proof. Let $(G^{\mathrm{T}}G)^{-1}\, G^{\mathrm{T}} \equiv (\gamma_{ij})$. From the proof of Theorem A.1 we have, for $i = 1, 2,..., n, j = 1, 2,..., n + 1$,

$$\begin{aligned} \gamma_{ij} &= \Big(\sum_{k\neq i} D_k^2\Big)\Big/(\Delta_2 g_i), \qquad & i = j, \\ &= (-1)^{i+j-1}\, D_i D_j/(\Delta_2 g_i), \qquad & i \neq j. \end{aligned} \tag{2.13}$$

(i) It follows from (2.13) and Lemma A.1 that $\sigma = 0$ implies

$$\sum_{j=1}^{n+1} \gamma_{ij} b_j = b_i/g_i\,, \qquad i = 1, 2,..., n.$$

(ii) From (2.9a), we have

$$\sum_{j=1}^{n+1} \gamma_{ij} b_j^{(p)} = (1/\Delta_q) \sum_{k=1}^{n+1} |\, D_k\,|^q\, c_{ik}\,, \qquad i = 1, 2,..., n,\ 1 < p \leqslant \infty,$$

where

$$c_{ik} = \sum_{j=1}^{n+1} \gamma_{ij} b_j + (\sigma(-1)^k \gamma_{ik}/D_k).$$

From (2.13), the above equation yields

$$\begin{aligned} c_{ik} &= \frac{1}{D_k g_k} \sum_{j \neq k} (-1)^{k+j-1} b_j D_j, \qquad & i = k, \\ &= b_i/g_i, & i \neq k. \end{aligned}$$

Hence $c_{ik} = w_i^k$ (cf. (2.12)). This completes the proof for the case $1 < p \leqslant \infty$, while the result for $p = 1$ follows by letting $p \to 1^+$.

THEOREM 2.2. *Let $A \in \mathscr{H}$ and, for each $k = 1, 2,..., n + 1$, let Z^k be the unique solution of $A^k Z^k = b^k$.*

(i) *If $\sigma = 0$, then $Ax = b$ is solvable and it has the unique solution $x = Z^k$, $k = 1, 2,..., n + 1$ (all the Z^k's being equal).*

(ii) *If $\sigma \neq 0$, then*

$$\xi_p = (1/\Delta_q) \sum_{k=1}^{n+1} | D_k |^q Z^k \tag{2.14}$$

is the unique l^p-solution to $Ax = b$ for $1 < p \leqslant \infty$. If λ and τ are defined as in Theorem 2.1, then

$$\xi_1 = (1/\tau) \sum_{k \in \lambda} Z^k \tag{2.15}$$

is an l^1-solution to $Ax = b$ which is unique if and only if λ is a singleton set ($\tau = 1$). Furthermore, the l^p-solution provided here is continuous with respect to p for $1 \leqslant p \leqslant \infty$.

Proof. Since A^k is nonsingular, it follows from $A^k P = G^k$, (2.10), and (2.11) that

$$Z^k = PW^k, \qquad k = 1, 2,..., n + 1. \tag{2.16}$$

(i) From (2.12), $\sigma = 0$ implies $w_i^k = b_i/g_i$, $i = 1, 2,..., n$, which is independent of k. Hence $W^k = W \equiv (b_i/g_i)$, $k = 1, 2,..., n + 1$, and $Z^k = PW^k = PW$, $k = 1, 2,..., n + 1$. From Theorem 2.1, $Ax = b$ is uniquely solvable with solution $x = (A^{\mathrm{T}}A)^{-1} A^{\mathrm{T}}b$. Therefore (A.1) and Lemma 2.1 show that $x = PW = Z^k$.

(ii) From (2.5) and (A.1) we have

$$\xi_p = P(G^TG)^{-1} G^T b^{(p)}. \tag{2.17}$$

Hence (2.14) and (2.15) follow from (2.17), Lemma 2.1, and (2.16).

Remark. Theorem 2.2 demonstrates how an l^p-solution of an overdetermined system with $A \in \mathscr{H}$ may be expressed as a convex combination of the solutions of the $n \times n$ subsystems.

3. A Stochastic Overdetermined System

Consider an overdetermined system $Ax = b$, where $A \in \mathscr{H}$ is a constant matrix but now $b = (b_i)$ is a random vector. Let $E(\cdot)$, $V(\cdot)$, and $\mathrm{Cov}(\cdot, \cdot)$ denote the expected value, variance, and covariance operators, respectively. We assume $V(b_i) < \infty$ for $i = 1, 2,..., n + 1$. From Theorem 2.1 we have

$$\xi_i^p = \sum_{j=1}^{n+1} \beta_{ij}(p)\, b_j, \qquad i = 1,..., n,\ 1 \leqslant p \leqslant \infty, \tag{3.1}$$

and hence

$$E(\xi_i^p) = \sum_{j=2}^{n+1} \beta_{ij}(p)\, E(b_j), \qquad i = 1,..., n,\ 1 \leqslant p \leqslant \infty. \tag{3.2}$$

We note that $E(\xi_p)$ is the l^p-solution to $Ax = E(b)$. We call $E(\xi_{\hat{p}})$ our "preferred approximate solution" to the stochastic overdetermined system if $\hat{p}$ is selected according to the minimum variance condition

$$v(\hat{p}) = \min_{1 \leqslant p \leqslant \infty} v(p),$$

where

$$v(p) = \sum_{i=1}^{n} V(\xi_i^p)$$

is the trace of the covariance matrix. Note that $v(p)$ is invariant under orthogonal transformations and is equal to the sum of the eigenvalues. Then, in a sense, it measures the total spread of the random variables. If $| D_i | = c$ for $i = 1, 2,..., n + 1$ then by Corollary 2.1 $v(p)$ is independent of p and thus any value of p, $1 \leqslant p \leqslant \infty$ will suffice for $\hat{p}$. Applying the variance operator to (3.1), we obtain

$$V(\xi_i^p) = \sum_{j=1}^{n+1} \beta_{ij}^2(p)\, V(b_j) + 2 \sum_{j<k} \beta_{ij}(p)\, \beta_{ik}(p)\, \mathrm{Cov}(b_j, b_k).$$

Therefore

$$v(p) = \sum_{j=1}^{n+1} V(b_j) \sum_{i=1}^{n} \beta_{ij}^2(p) + 2 \sum_{j<k} \operatorname{Cov}(b_j, b_k) \sum_{i=1}^{n} \beta_{ij}(p)\, \beta_{ik}(p). \tag{3.3}$$

We note that $v(p)$ is continuous for $1 \leqslant p \leqslant \infty$ and differentiable for $1 < p < \infty$.

Without loss of generality, we then assume that $| a_{11} | > | a_{21} |$ for the case $n = 1$. Here

$$M = (m_{11}, m_{12}) = \left(\frac{a_{11}}{a_{11}^2 + a_{21}^2}, \frac{a_{21}}{a_{11}^2 + a_{21}^2} \right),$$

$$\begin{aligned} \beta_{1j}(p) &= \frac{| a_{j1} |^q}{a_{j1}(| a_{11} |^q + | a_{21} |^q)}, && 1 < p \leqslant \infty, \\ &= (1 + (-1)^{j+1})/2a_{11}, && p = 1, \end{aligned} \qquad j = 1, 2,$$

and from (3.2) and (3.3)

$$\begin{aligned} E(\xi_p) &= \frac{| a_{11} |^q \dfrac{E(b_1)}{a_{11}} + | a_{21} |^q \dfrac{E(b_2)}{a_{21}}}{| a_{11} |^q + | a_{21} |^q}, && 1 < p \leqslant \infty, \\ &= E(b_1)/a_{11}, && p = 1, \\ v(p) &= \frac{K_1}{(| a_{11} |^q + | a_{21} |^q)^2}, && 1 < p \leqslant \infty, \\ &= V(b_1)/a_{11}^2, && p = 1, \end{aligned}$$

where

$$K_1 = | a_{11} |^{2q-2} V(b_1) + | a_{21} |^{2q-2} V(b_2) + 2 \frac{| a_{11}a_{21} |^q}{a_{11}a_{21}} \operatorname{Cov}(b_1, b_2).$$

If $V(b_1) = 0$ and $V(b_2) \neq 0$ we choose $\hat{p} = 1$, while if $V(b_1) \neq 0$ and $V(b_2) = 0$ we choose $\hat{p} = \infty$. For $V(b_1) \neq 0$ and $V(b_2) \neq 0$ we find $\lim_{p \to 1^+} (dv/dp) = \lim_{p \to \infty}(dv/dp) = 0$. Furthermore, $dv/dp = 0$ at $p = p_0$, where

$$\begin{aligned} p_0 &= 1, && \theta = 0, \\ &= \infty, && \theta = 1, \\ &= 1 + \frac{\ln(| a_{11} |/| a_{21} |)}{\ln | \theta |}, && \text{otherwise}, \end{aligned} \tag{3.4}$$

with

$$\theta = \frac{|a_{11}|[V(b_2) - (a_{21}/a_{11}) \operatorname{Cov}(b_1, b_2)]}{|a_{21}|[V(b_1) - (a_{11}/a_{21}) \operatorname{Cov}(b_1, b_2)]}.$$

Now $p_0 \geqslant 1$ if and only if $\theta \geqslant 1$, in which case

$$v(p_0) = \frac{V(b_1) V(b_2) - \operatorname{Cov}(b_1, b_2)}{a_{11}^2 V(b_2) + a_{21}^2 V(b_1) - 2a_{11}a_{21} \operatorname{Cov}(b_1, b_2)}.$$

If $\theta \geqslant 1$, we choose $\hat{p}$ such that $v(\hat{p}) = \min\{v(1), v(p_0), v(\infty)\}$; otherwise, we choose $\hat{p}$ such that $v(\hat{p}) = \min\{v(1), v(\infty)\}$. If, in addition, b_1 and b_2 are uncorrelated, then $\theta = (|a_{11}| V(b_2))/(|a_{21}| V(b_1))$ and we find explicitly that

$$\begin{aligned} \hat{p} &= \infty & 0 < \theta \leqslant 1 \\ &= 1 + \frac{\ln(|a_{11}|/|a_{21}|)}{\ln|\theta|} & \theta > 1. \end{aligned}$$

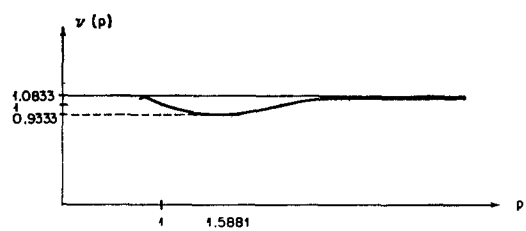
We now present two examples, the solutions of which were obtained utilizing a computer. $v(p)$ was calculated according to (3.3), thus deciding $\hat{p}$. The corresponding graph was also provided. Note that any type of distribution with $V(b_i) < \infty$ for $i = 1, 2, \ldots, n + 1$ and yielding the same first two moments would produce the same choice of $\hat{p}$.

EXAMPLE 1. $A = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and the joint distribution of b is

$b_1 \backslash b_2$	1	5
2	$\frac{3}{8}$	$\frac{1}{8}$
6	$\frac{1}{4}$	$\frac{1}{4}$

Hence $E(b_1) = 4$, $E(b_2) = 2.5$, $V(b_1) = 4$, $V(b_2) = 3.75$, $\operatorname{Cov}(b_1, b_2) = 1$. Here $\theta > 1$ and therefore (3.4) yields $p_0 = 1.5881$. Since $v(1) = 1$, $v(p_0) = 0.9333$ and $v(\infty) = 1.0833$, we choose $\hat{p} = 1.5881$. Hence our preferred approximate solution is

$$E(\xi_{1.5881}) = 2.0666,$$



EXAMPLE 2.

$$A = \begin{pmatrix} 0.5 & 1 & 0.5 \\ -3 & -9 & -6 \\ 4 & 12 & 6 \\ 2.5 & 6.5 & 3.5 \end{pmatrix}$$

and the joint distribution of b is $b_4 = 0$,

$b_3 = 4$

$b_1 \backslash b_2$	1	5
2	1/8	1/4
6	1/8	0

$b_3 = 6$

$b_1 \backslash b_2$	1	5
2	1/16	1/16
6	1/8	1/4

Thus

$$E(b_1) = 4, \quad E(b_2) = 3.25, \quad E(b_3) = 5, \quad E(b_4) = 0,$$

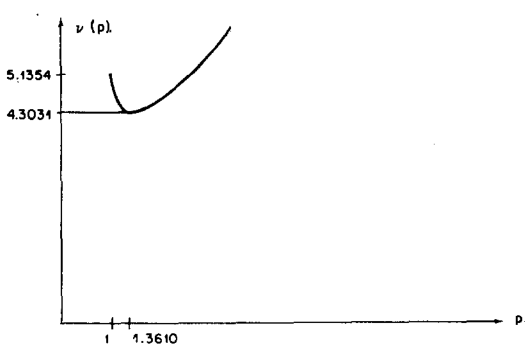
$$V(b_1) = 4, \quad V(b_2) = 3.9375, \quad V(b_3) = 1, \quad V(b_4) = 0,$$

$$\mathrm{Cov}(b_1, b_2) = -0.5, \quad \mathrm{Cov}(b_1, b_3) = 1, \quad \mathrm{Cov}(b_2, b_3) = 0.25,$$

$$\mathrm{Cov}(b_1, b_4) = \mathrm{Cov}(b_2, b_4) = \mathrm{Cov}(b_3, b_4) = 0.$$

Here $\hat{p} = 1.3610$,

$$E(\xi_{1.3610}) = \begin{pmatrix} -2.8273 \\ 3.6896 \\ -4.6631 \end{pmatrix},$$



APPENDIX: Some Results on Matrices

As a consequence of the Binet–Cauchy theorem [3], we have

Lemma A.1. *If A is an $(n+1) \times n$ matrix, then* $\det(A^{\mathrm{T}}A) = \Delta_2$. *We note that for $A \in \mathscr{H}$, $\Delta_2 \neq 0$.*

Theorem A.1. *If $A \in \mathscr{H}$, then*

$$I - A(A^{\mathrm{T}}A)^{-1} A^{\mathrm{T}}$$
$$= \frac{1}{\Delta_2} \begin{pmatrix} D_1{}^2 & -D_1D_2 & \cdots & (-1)^n D_1D_{n+1} \\ -D_2D_1 & D_2{}^2 & \cdots & (-1)^{n+1} D_2D_{n+1} \\ \vdots & \vdots & & \vdots \\ (-1)^n D_{n+1}D_1 & (-1)^{n+1} D_{n+1}D_2 & \cdots & (-1)^{2n} D_{n+1}^2 \end{pmatrix}$$
$$= \frac{\delta\delta^{\mathrm{T}}}{\Delta_2}.$$

Proof. Since $AP = G$, it follows that

$$(A^{\mathrm{T}}A)^{-1} A^{\mathrm{T}} = P(G^{\mathrm{T}}G)^{-1} G^{\mathrm{T}} \tag{A.1}$$

and

$$A(A^{\mathrm{T}}A)^{-1} A^{\mathrm{T}} = G(G^{\mathrm{T}}G)^{-1} G^{\mathrm{T}}. \tag{A.2}$$

From the definition of G^k, we see that

$$\begin{aligned} \hat{D}_k &= (-1)^{n-k} h_k \prod_{i \neq k} g_i, \qquad & k &= 1, 2, \ldots, n, \\ &= \prod_{i=1}^{n} g_i, & k &= n+1. \end{aligned}$$

Since $\hat{D}_k = D_k \det(P)$ and $|\det(P)| = 1$, we have $D_iD_j = \hat{D}_i\hat{D}_j$ for $i, j = 1, 2, \ldots, n+1$. From Lemma A.1,

$$\det(A^{\mathrm{T}}A) = \sum_{k=1}^{n+1} D_k{}^2 = \sum_{k=1}^{n+1} \hat{D}_k{}^2 = \sum_{k=1}^{n} h_k{}^2 \prod_{i \neq k} g_i{}^2 + \prod_{i=1}^{n} g_i{}^2.$$

Since

$$G^{\mathrm{T}}G = \begin{pmatrix} g_1{}^2 + h_1{}^2 & h_1h_2 & \cdots & h_1h_n \\ h_2h_1 & g_2{}^2 + h_2{}^2 & \cdots & h_2h_n \\ \vdots & \vdots & & \vdots \\ h_nh_1 & h_nh_2 & \cdots & g_n{}^2 + h_n{}^2 \end{pmatrix},$$

we have

$$(G^{\mathrm{T}}G)^{-1} =$$

$$\frac{1}{\Delta_2}\begin{bmatrix} \Big(\sum_{k\neq 1}\frac{h_k^2}{g_n^2}+1\Big)\prod_{i\neq 1} g_i^2 & -h_1h_2\prod_{i\neq 1,2} g_i^2 & \cdots & -h_1h_n\prod_{i\neq 1,n} g_i^2 \\ -h_1h_2\prod_{i\neq 2,1} g_i^2 & \Big(\sum_{k\neq 2}\frac{h_k^2}{g_k^2}+1\Big)\prod_{i\neq 2} g_i^2 & \cdots & -h_2h_n\prod_{i\neq 2,n} g_i^2 \\ \vdots & \vdots & & \vdots \\ -h_nh_1\prod_{i\neq n,1} g_i^2 & -h_nh_2\prod_{i\neq n,2} g_i^2 & \cdots & \Big(\sum_{k\neq n}\frac{h_k^2}{g_k^2}+1\Big)\prod_{i\neq n} g_i^2 \end{bmatrix}.$$

Hence

$$G(G^{\mathrm{T}}G)^{-1}\, G^{\mathrm{T}} =$$

$$\frac{1}{\Delta_2}\begin{bmatrix} \sum_{k\neq 1} D_k^2 & D_1D_2 & -D_1D_3 & \cdots\ (-1)^{n+1}D_1D_{n+1} \\ D_2D_1 & \sum_{k\neq 2} D_k^2 & D_2D_3 & \cdots\ (-1)^{n+2}D_2D_{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ (-1)^{n+1}D_{n+1}D_1 & (-1)^{n+2}D_{n+1}D_2 & (-1)^{n+3}D_{n+1}D_3\ \cdots & \sum_{k\neq n+1} D_k^2 \end{bmatrix}.$$

The result now follows in view of Lemma A.1 and (A.2).

LEMMA A.2. *If* $A \in \mathscr{H}$ *and* M_1 *and* M_2 *are both* $n \times r$ *matrices then* $AM_1 = AM_2$ *implies that* $M_1 = M_2$.

THEOREM A.2. *If* $A \in \mathscr{H}$, *then* $(A^{\mathrm{T}}A)^{-1}\, A^{\mathrm{T}} = M = (m_{ij})$, *where* m_{ij} *is defined in* (2.6).

Proof. By Lemma A.2, this theorem is proved if $AM = A(A^{\mathrm{T}}A)^{-1}\, A^{\mathrm{T}}$. The entry of AM at the lth row and the jth column is

$$\sum_{i=1}^{n} a_{li}m_{ij} = \frac{1}{\Delta_2}\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} & (-1)^n\, D_1 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{j-1,1} & a_{j-1,2} & \cdots & a_{j-1,n} & (-1)^{n-j+2}\, D_{j-1} \\ a_{l,1} & a_{l,2} & \cdots & a_{l,n} & 0 \\ a_{j+1,1} & a_{j+1,2} & \cdots & a_{j+1,n} & (-1)^{n-1}\, D_{j+1} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{n+1,1} & a_{n+1,2} & \cdots & a_{n+1,n} & (-1)^0\, D_{n+1} \end{vmatrix},$$

where $l = 1,..., n + 1, j = 1,..., n + 1$.

If $l = j$, the determinant expanded about the last column yields

$$\sum_{i=1}^{n} a_{ji} m_{ij} = (1/\Delta_2) \sum_{i \neq j} D_i^2.$$

If $l \neq j$, the same expansion yields zero for each entry except the lth one, for which we have

$$\sum_{i=1}^{n} a_{li} m_{ij} = ((-1)^{l+j-1}/\Delta_2)\, D_l D_j\,.$$

Therefore, we see that matrix AM is precisely $G(G^{\mathrm{T}}G)^{-1}\, G^{\mathrm{T}} = A(A^{\mathrm{T}}A)^{-1}\, A^{\mathrm{T}}$ as constructed in Theorem A.1.

References

1. C. S. Duris, "An Algorithm for Solving Overdetermined Linear Equations in the l^p-Space," Math. Report 70–10, Drexel University, 1970.
2. C. S. Duris and V. P. Sreedharan, Chebyshev and l^1 solutions of linear equations using least squares solutions, *SIAM J. Numer. Anal.* **5** (1968), 491–505.
3. M. Marcus, "Basic Theorems in Matrix Theory," National Bureau of Standards, Applied Mathematics Series 57, 1960.
4. V. P. Sreedharan, Solution of overdetermined linear equations which minimize error in an abstract norm, *Numer. Math.* **13** (1969), 146–151.